# INVESTIGATING THE IMPACT OF HIGH FRAME RATES ON VIDEO COMPRESSION

*Alex Mackin, Fan Zhang, Miltiadis Alexios Papadopoulos and David Bull*

Department of Electrical and Electronic Engineering, University of Bristol, BS8 1UB, UK
{A.Mackin, Fan.Zhang, MA.Papadopoulos, Dave.Bull}@bristol.ac.uk

## ABSTRACT

In this paper we investigate the impact of frame rate variation on HEVC video compression, and demonstrate that high frame rates (60+ fps) can lead to increased perceptual quality, notably in high bitrate environments. In order to quantify content dependence, a novel way of partitioning video sequences into categories is proposed. Results show that rate-quality performance is improved at higher frame rates for video sequences with camera motion, whereas lower frame rates are favorable in sequences with complex motion (e.g. dynamic textures). We calculate that 60 fps and 120 fps are optimal choices of frame rate at bitrates of 3 Mbps and 7 Mbps respectively, demonstrating that increased frame rates are both feasible and desirable, given current broadcast data rates.

***Index Terms***— High frame rates, HFR, HEVC, Video Compression, Immersive Video

## 1. INTRODUCTION

Frame rates have remained static for a number of years, and in the case of cinema, are unchanged since the 1920's [1]. The most recent UHDTV standard ITU-R BT.2020-2 [2] supports extended video parameters compared to its predecessor [3], including frame rates up to 120 fps. However, frame rates in current UHDTV formats rarely exceed 60 fps, while dynamic range, bit-depth and spatial resolution are all extended.

Higher frame rates have though recently stimulated interest in the broadcast [4, 5] and virtual reality [6] communities, due to a number of clear benefits: the visibility of temporal aliasing artefacts is diminished [7–12]; there is a reduction in perceptible motion blur [10–14]; viewer stress levels are reduced (signified by a lower blinking frequency) [15]; increased realism, smoother motion, improved depth perception for both expert [16] and non-expert [17] viewers; and an increase in perceptual quality [18], at least up to 240 fps [19].

To assess the feasibility of high frame rates, a rate-quality analysis scrutinizing the role of frame rate in video compression is required. In previous work, low spatial (CIF) and temporal resolutions (up to 30 fps), and outdated compression standards (H.264/AVC), were considered [20, 21]. A frame rate selection method up to 60 fps was recently presented [22].

This paper investigates for the first time, the influence of frame rate (up to 120 fps) on HEVC video compression. All the video sequences in the recently published BVI-HFR video database [18] were encoded, and subsequent analysis demonstrates that the clear perceptual benefits associated with high frame rates are accessible at current data rates. Our results establish that content dependency related to motion exists.

## 2. METHODOLOGY

The BVI-HFR video database [18] contains 22 unique uncompressed video sequences at HD resolution (1920×1080) and 120 fps. Each video sequence has further been temporally down-sampled by averaging frames to 60, 30 and 15 fps - resulting in a total of 88 sequences. Subjective evaluations (in the form of MOS scores) are provided for each sequence.

The middle three-seconds of each of the 88 sequences was encoded using the HEVC [23] reference codec (HM 16.4) at five quantization parameters (QP): 22, 27, 32, 37, 42; and using three common compression modes: All Intra, Low Delay and Random Access [24] (1320 encoded sequences in total).

The SQF quality metric [25] has been validated on data that contains variations in frame rate and QP levels, and is used here to predict the quality of the compressed sequences:

$$\mathrm{SQF} = \hat{Q}\left(1 - 1 \,/\, \left[1 + e^{\hat{p}(Q-\hat{s})}\right]\right) \tag{1}$$

where $\hat{Q}$ is the MOS score of the uncompressed sequence, $Q$ is the PSNR of the compressed sequence, and $\hat{s}, \hat{p}$ are model parameters. The parameter $\hat{s}$ is calculated using a linear combination of features. A value of $\hat{p} = 0.34$ is suggested [25].

As to reduce time complexity, only the middle 3s of each sequence (originally 10 seconds in length) was encoded. Moss *et al.* [26] have shown similarity in MOS scores between these two video lengths - assuming that the sequences are temporally consistent. A Mann-Whitney $U$ test[1] reports no significant ($p$ < 0.05) difference in temporal information [27] ($U$ = 3546, $p$ = 0.34), motion activity intensity [25] ($U$ = 3565, $p$ = 0.36) and motion direction activity [25] ($U$ = 3378, $p$ = 0.14) between the middle 3s of the video sequences and the remaining 7s (thus ensuring independent samples). These results validate our approach of using the MOS scores from the 10s video sequences for $\hat{Q}$ in Eq. 1.

The authors acknowledge BBC R&D and EPSRC for funding through EPSRC iCASE award 12440623 and grant EP/M000885/1 respectively.

[1] As normality cannot be assumed (verified by the Kolmogorov-Smirnov test), an independent samples t-test nor a multivariate ANOVA can be used.

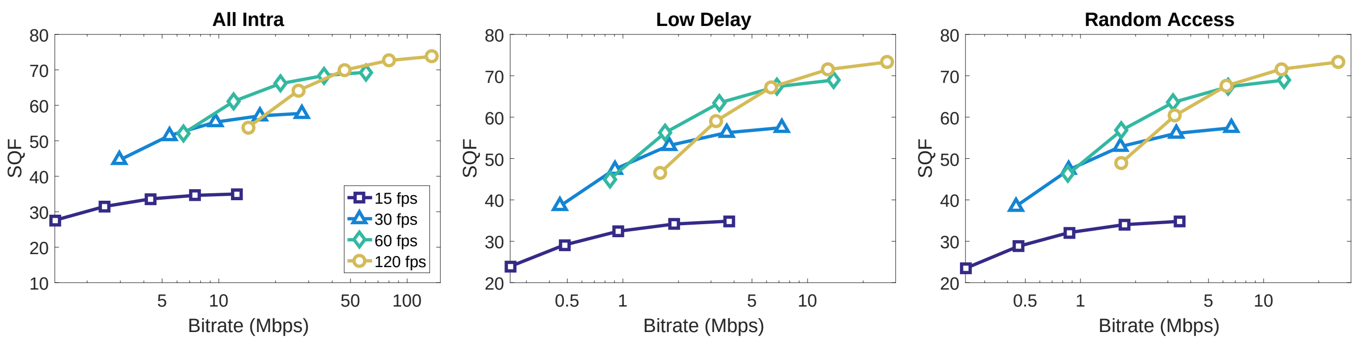


**Fig. 1**: Rate-quality curves for the tested frame rates and compression modes, showing the average over all 22 sequences in the BVI-HFR video database.

## 3. RESULTS AND ANALYSIS

### 3.1. Overall Performance

Table 1 reports the internal distribution of bits by the HM encoder per frame, over the range of tested frame rates. For All Intra mode, the number of bits increases in all areas with frame rate, suggesting that the increased spatial complexity associated with increased frame rates is harder to encode (due to a reduction in motion blur [18]). For Low Delay and Random Access modes, the use of motion prediction dramatically decreases the number of bits (as intra coding generally consumes less bits than inter coding). At higher frame rates, the increased temporal correlation between frames leads to smaller motion vectors that more accurately represent motion in the scene [18] - signified by the reduction in bits consumed by coding residuals as frame rates increase.

**Table 1**: Average number of bits (kb) consumed by the HM encoder per frame. R = Residual Coding, MP = Motion Prediction, I = Intra Direction, MI = Merge Index, MS = Mode Signaling, P = Partitioning and O = Other.

| Frame Rate | R | MP | I | MI | MS | P | O |
|---|---|---|---|---|---|---|---|
| **All Intra** | | | | | | | |
| 15 | 345 | - | 24 | - | - | 6 | 1 |
| 30 | 378 | - | 27 | - | - | 7 | 1 |
| 60 | 415 | - | 30 | - | - | 8 | 2 |
| 120 | 459 | - | 34 | - | - | 9 | 2 |
| **Low Delay** | | | | | | | |
| 15 | 81 | 5 | 3 | 2 | 3 | 3 | 0 |
| 30 | 76 | 5 | 3 | 3 | 3 | 4 | 1 |
| 60 | 70 | 5 | 3 | 3 | 3 | 4 | 1 |
| 120 | 67 | 4 | 3 | 3 | 3 | 4 | 1 |
| **Random Access** | | | | | | | |
| 15 | 75 | 4 | 4 | 2 | 2 | 3 | 0 |
| 30 | 70 | 5 | 4 | 2 | 2 | 3 | 0 |
| 60 | 67 | 4 | 4 | 2 | 2 | 3 | 0 |
| 120 | 66 | 4 | 4 | 2 | 2 | 3 | 1 |

Fig. 1 shows rate-quality curves for all tested frame rates, QP values and compression modes. Each data point is the average of the 22 unique sequences in the BVI-HFR video database. Higher frame rates exhibit improved rate-quality performance at high bitrates and low QP values for all modes.

The rate-quality curves in Fig. 1 can be used to predict frame rates which maximize perceptual quality as a function of bitrate. This is achieved by fitting an exponential curve to the rate-quality data of each sequence (between the minimum and maximum bitrates for all frame rates). (Pareto-) Optimal frame rates then lie on the convex hull of these fitting curves. We define a *transition point* as the bitrate at which the optimal frame rate changes. Fig. 2 shows the distribution of these transition points, and results indicate a quasi-linear relationship between average bitrate and the optimal frame rate.

### 3.2. Quantifying Content Dependence

The spread of the transition points in Fig. 2 is fairly large, indicating as expected, a degree of content dependence in optimal frame rate selection. Ma *et al.* [20] propose a method for relating the source statistics of sequences with frame rate variations to rate-quality performance, and therefore optimal frame rates. However during informal testing on the BVI-HFR video database, unexpected and inconsistent results were observed - assumed to be due to the model being based on H.264/AVC and low frame rates up to 30 fps. Therefore in order to quantify content dependence, we propose a novel way to separate the sequences up into different categories.

**Table 2**: The proposed groupings of the sequences in the BVI-HFR database, where **bold** indicates camera motion. DFD values are shown in brackets.

| Motion | Sequences |
|---|---|
| Simple | bobblehead (2), bouncyball (2.7), catch (2.7), flowers (3.7), hamster (5), golf_side (2.6), guitar_focus (2.5), martial_arts (1.3), pond (2.5), pour (5.1), typing (2.4), **books (4.8)**, **catch_track (3.6)**, **cyclist (5.1)**, **joggers (5.7)**, **library (2.3)** |
| Complex | leaves_wall (7), lamppost (12.4), plasma (6.6), sparkler (13.4), water_ripples (7.2), water_splashing (14.3) |

Increased frame rates lead to clear perceptual benefits in video sequences that contain camera motion [18], while sequences with complex motion (e.g. dynamic textures) are generally more difficult to encode [28], due to (in part) the use of linear motion vectors, the increased use of intra blocks, and the influence of finer block partition in intra mode.

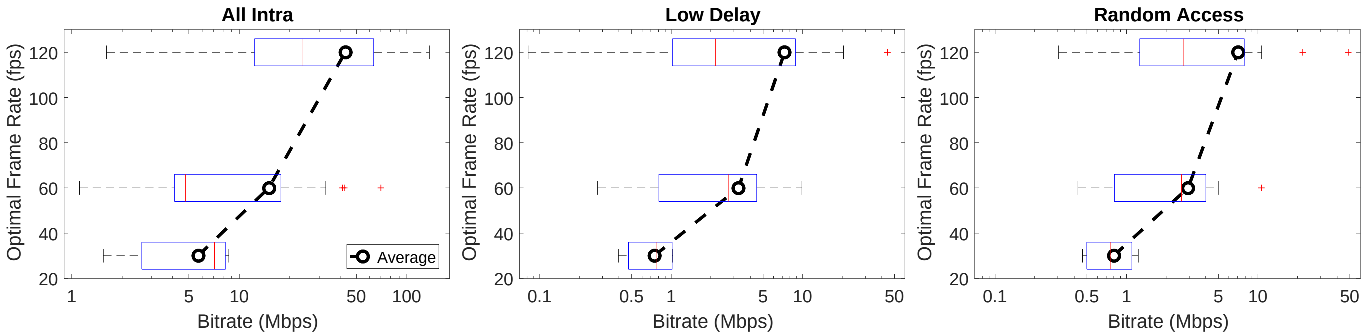


**Fig. 2**: Boxplots showing the distribution of transition points (the $y$-axis is the optimal frame rate that we change to at the transition point). The (blue) box shows the interquartile range (IQR) of the data, the whiskers are 1.5 IQR, the vertical (red) line is the median, and the (black) dashed line is the mean.

Therefore we propose grouping the 22 source sequences (120 fps) in the BVI-HFR video database into those with (5) and without (17) camera motion, and into those with simple (16) and complex (6) motion (see Table 2). Complexity of motion is quantified here using the displaced frame difference (DFD) feature [25] (based on motion estimation[2]). We interpret that a sequence contains complex motion if DFD $\geq 6.5$.

**Table 3**: Average ($\mu_B$) and average difference ($\delta_B$) in bit allocation per frame by the HM encoder (kb) between the tested frame rates (rounded).

| Category | $\mu_B/\delta_B$ | R | MP | I | MI | MS | P | O |
|---|---|---|---|---|---|---|---|---|
| **All Intra** | | | | | | | | |
| No Camera | $\mu_B$ | 439 | 0 | 32 | 0 | 0 | 8 | 2 |
| | $\delta_B$ | 27 | 0 | 3 | 0 | 0 | 1 | 0 |
| Camera | $\mu_B$ | 264 | 0 | 18 | 0 | 0 | 5 | 1 |
| | $\delta_B$ | 77 | 0 | 4 | 0 | 0 | 1 | 0 |
| Simple | $\mu_B$ | 370 | 0 | 28 | 0 | 0 | 7 | 2 |
| | $\delta_B$ | 34 | 0 | 3 | 0 | 0 | 1 | 0 |
| Complex | $\mu_B$ | 478 | 0 | 31 | 0 | 0 | 8 | 1 |
| | $\delta_B$ | 50 | 0 | 6 | 0 | 0 | 1 | 0 |
| **Low Delay** | | | | | | | | |
| No Camera | $\mu_B$ | 84 | 5 | 3 | 3 | 3 | 4 | 1 |
| | $\delta_B$ | -3 | -0 | -0 | 0 | 0 | 0 | 0 |
| Camera | $\mu_B$ | 38 | 2 | 1 | 2 | 2 | 2 | 0 |
| | $\delta_B$ | -10 | 0 | -1 | 0 | 0 | 0 | 0 |
| Simple | $\mu_B$ | 38 | 3 | 2 | 2 | 2 | 3 | 0 |
| | $\delta_B$ | -7 | -0 | -1 | 0 | -0 | -0 | 0 |
| Complex | $\mu_B$ | 170 | 9 | 6 | 4 | 5 | 7 | 1 |
| | $\delta_B$ | 2 | -0 | 1 | 1 | 1 | 1 | 0 |
| **Random Access** | | | | | | | | |
| No Camera | $\mu_B$ | 79 | 5 | 4 | 2 | 3 | 4 | 1 |
| | $\delta_B$ | -2 | -0 | 0 | 0 | 0 | 0 | 0 |
| Camera | $\mu_B$ | 37 | 2 | 2 | 1 | 1 | 2 | 0 |
| | $\delta_B$ | -6 | 0 | -1 | 0 | -0 | 0 | 0 |
| Simple | $\mu_B$ | 41 | 3 | 3 | 1 | 2 | 2 | 0 |
| | $\delta_B$ | -5 | -0 | -0 | 0 | -0 | -0 | 0 |
| Complex | $\mu_B$ | 147 | 8 | 7 | 3 | 4 | 5 | 1 |
| | $\delta_B$ | 2 | 0 | 1 | 0 | 0 | 1 | 0 |

[2] With an 8×8 block size, exhaustive search and a search range of 64.

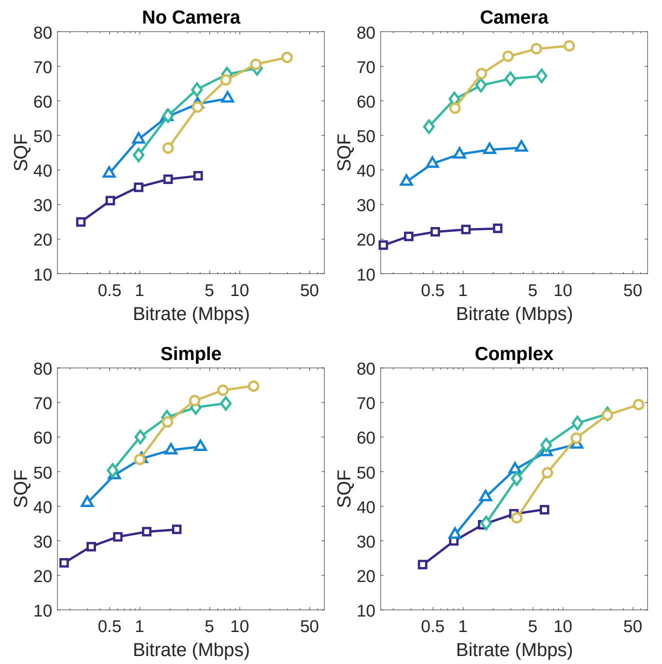


**Fig. 3**: Rate-quality curves for the video groupings in Random Access mode.

Table 3 reports the average ($\mu_B$) and average difference ($\delta_B$) in bit allocation per frame between the tested frame rates, for the proposed video groupings. $\mu_B$ and $\delta_B$ are defined as:

$$\begin{aligned} \mu_B &= (B_{15} + B_{30} + B_{60} + B_{120})/4 \\ \delta_B &= (B_{120} - B_{15})/3 \end{aligned} \tag{2}$$

where $B_F$ is the number of bits allocated at frame rate $F$.

For all modes, the average number of bits is higher in video sequences containing complex compared to simple motion, and no camera compared to camera motion. For All Intra Mode, video sequences containing camera motion have considerably fewer bits allocated in all areas than the other groupings, suggesting that the associated increase in motion blur is easier to encode (smaller valued high frequency DCT coefficients). For Low Delay and Random Access modes, the number of bits consumed during residual coding decreases as frame rates increase, except for the case of complex motion.

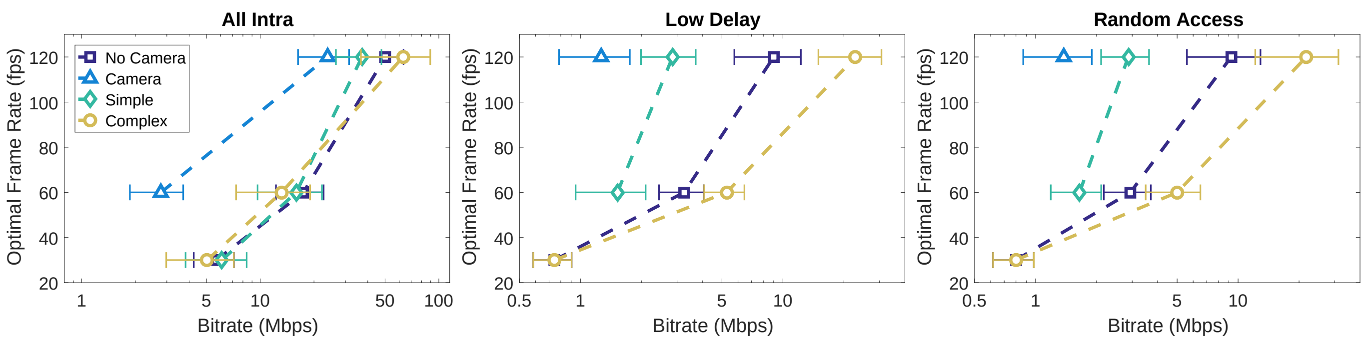


**Fig. 4**: A comparison of the average transition points between the video groupings. Error bars represent standard error of the mean.

Fig. 3 shows rate-quality curves for each of the video groupings in Random Access mode (Low Delay shows similar results). Video sequences with camera motion - and to a lesser extent simple motion - show increased rate-quality performance at the higher frame rates tested. On the other hand, video sequences with no camera or complex motion show improved rate-quality performance at the lower frame rates.

Optimal frame rates can be calculated from this rate-quality data (using the method in Section 3.1). Fig. 4 shows a comparison between the average transition points of the video groupings, and results demonstrate clear content dependency. For all modes, video sequences with camera motion have higher optimal frame rates at all tested bitrates.

Video sequences containing complex motion have lower optimal frame rates compared to the other groupings (expect for All Intra mode, as there is no motion prediction). This is postulated to be due to a lack of merge modes, and that current linear motion models cannot faithfully represent underlying non-linear motion. Therefore motion estimation will either be less accurate (higher DFD), or a higher proportion of coding units (CU) will be intra coded (see the increase in bits consumed during residual coding with frame rate in Table 3 for complex motion). For the case of static and dynamic textures, texture masking may conceal coding artifacts and the increased levels of temporal aliasing and motion blur associated with lower frame rates [11].

## 4. DISCUSSION

We have demonstrated that frame rates - higher than those conventionally used today (60+ fps) - can be beneficial even at relatively low data rates. However before high frame rates become prevalent, more scrutiny is needed to further exploit the source statistics during encoding. The relatively poor rate-quality performance of HEVC for sequences containing complex motion demonstrates that different motion models or coding modes may need to be considered, in an attempt to characterize underlying non-linear and nuanced motion. Rate-quality performance may further be improved by normalizing group of pictures (GOP) length e.g. a GOP length of 8 at 30 fps has the same temporal span as length 16 at 60 fps.

The reduction in motion blur associated with higher frame rates [18] leads to increased spatial complexity and subsequently higher valued high frequency DCT coefficients. Future intra coding methods should exploit this fact, as currently the number of bits per frame increases by approximately 10% when doubling frame rates in All Intra mode (Table 1).

The bitrates at which 60 and 120 fps become the optimal choice in frame rate is on average around 3 Mbps and 7 Mbps respectively for Low Delay and Random Access modes (from Fig. 4). The recommended bitrate to stream a HD resolution video on Netflix is 5 Mbps [29]. Approximately 14 of the 22 sequences (65%) in the BVI-HFR video database had an optimal frame of at least 60 fps at this bitrate (from Fig. 2).

The following table reports the distribution of optimal frame rates at 5 Mbps for the proposed video groupings:

| Group | 15 fps | 30 fps | 60 fps | 120 fps |
|---|---|---|---|---|
| No Camera | 0% | 12% | 47% | 41% |
| Camera | 0% | 0% | 0% | 100% |
| Simple | 0% | 0% | 25% | 75% |
| Complex | 0% | 33% | 67% | 0% |

These results demonstrate that high frame rates (60 fps+) can provide clear perceptual benefits at current data rates.

In order to account for content dependency, frame rates should ideally be selected in an adaptive manner given the source statistics of the video sequence. However, before adaptive formats can be considered, the interplay between video parameters (HDR, 4K etc.) needs to be investigated. This will form part of our future work.

## 5. CONCLUSIONS

In this paper we have shown that high frame rates (60+ fps) should be considered for future video formats, as the clear perceptual benefits associated with increased frame rates are accessible at current data rates. In studying the distribution of bits during HEVC compression, we have ascertained where the encoder exploits the source statistics of higher frame rates, and where improvements are required - notably related to the lack of merge modes, the motion model used and intra coding. Results show that higher frame rates are advantageous in sequences containing simple and/or camera motion.

## 6. REFERENCES


[1] D. Bull, *Communicating Pictures: A Course in Image and Video Coding*, p. 455, Elsevier, 2014.

[2] ITU-R Recommendation BT. 2020-2, "Parameter values for ultra-high definition television systems for production and international programme exchange," 2015.

[3] ITU-R Recommendation BT. 709-6, "Basic parameter values for the hdtv standard for the studio and for international programme exchange," 2015.

[4] M. Armstrong, D. Flynn, M. Hammond, S. Jolly, and R. Salmon, "High frame-rate television," *BBC Research & Development White Paper 169*, 2008.

[5] K. Noland, "The application of sampling theory to television frame rate requirements," *BBC Research & Development White Paper 282*, 2014.

[6] D. Lanman, H. Fuchs, M. Mine, I. McDowall, and M. Abrash, "Put on your 3D glasses now: The past, present, and future of virtual and augmented reality," in *ACM SIGGRAPH 2014 Courses*, 2014, pp. 12:1–12:173.

[7] D. Hoffman, V. Karasev, and M. Banks, "Temporal presentation protocols in stereoscopic displays: Flicker visibility, perceived motion, and perceived depth," *Journal of the Society for Information Display*, vol. 19, no. 3, pp. 271, 2011.

[8] A. Watson, "High frame rates and human vision: A view through the window of visibility," *SMPTE Motion Imaging Journal*, vol. 122, no. 2, pp. 18–32, 2013.

[9] A. Mackin, K. Noland, and D. Bull, "High frame rates and the visibility of motion artifacts," *SMPTE Motion Imaging Journal*, vol. 126, no. 5, 2017.

[10] S. Daly, N. Xu, J. Crenshaw, and V. Zunjarrao, "A psychophysical study exploring judder using fundamental signals and complex imagery," in *SMPTE Conferences*. Society of Motion Picture and Television Engineers, 2014, vol. 2014, pp. 1–14.

[11] A. Mackin, K. Noland, and D. Bull, "The visibility of motion artifacts and their effect on motion quality," in *Image Processing (ICIP), 2016 23rd IEEE International Conference on*, 2016.

[12] Y. Kuroki, T. Nishi, S. Kobayashi, H. Oyaizu, and S. Yoshimura, "A psychophysical study of improvements in motion-image quality by using high frame rates," *Journal of the Society for Information Display*, vol. 15, pp. 61–68, 2007.

[13] M. Sugawara, K. Omura, M. Emoto, and Y. Nojiri, "Temporal sampling parameters and motion portrayal of television," *SID Symposium Digest of Technical Papers*, vol. 40, no. 1, pp. 1200–1203, 2009.

[14] R. Selfridge, K. Noland, and M. Hansard, "Visibility of motion blur and strobing artefacts in video at 100 frames per second," in *Proceedings of the 13th European Conference on Visual Media Production (CVMP 2016)*. ACM, 2016, p. 3.

[15] B. Tag, J. Shimizu, C. Zhang, K. Kunze, N. Ohta, and K. Sugiura, "In the eye of the beholder: The impact of frame rate on human eye blink," in *Human Factors in Computing Systems, 2016 CHI Conference on*. ACM, 2016, pp. 2321–2327.

[16] R. Allison, L. Wilcox, R. Anthony, J. Helliker, and B. Dunk, "Paper: Expert viewers preferences for higher frame rate 3D film," *Journal of Imaging Science and Technology*, vol. 60, no. 6, pp. 60402–1, 2016.

[17] L. Wilcox, R. Allison, J. Helliker, B. Dunk, and R. Anthony, "Evidence that viewers prefer higher frame-rate film," *ACM Transactions on Applied Perception (TAP)*, vol. 12, no. 4, pp. 15, 2015.

[18] A. Mackin, F. Zhang, and D. Bull, "A study of subjective video quality at various frame rates," in *Image Processing (ICIP), 2015 22nd IEEE International Conference on*, 2015.

[19] M. Emoto, Y. Kusakabe, and M. Sugawara, "High-frame-rate motion picture quality and its independence of viewing distance," *Journal of Display Technology*, vol. 10, no. 8, pp. 635–641, 2014.

[20] Z. Ma, M. Xu, Y. Ou, and Y. Wang, "Modeling of rate and perceptual quality of compressed video as functions of frame rate and quantization stepsize and its applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 5, pp. 671–682, 2012.

[21] Y. Ou, Y. Xue, and Y. Wang, "Q-STAR: A perceptual video quality model considering impact of spatial, temporal, and amplitude resolutions," *IEEE Transactions on Image Processing*, vol. 23, no. 6, pp. 2473–2486, June 2014.

[22] Q. Huang, S. Jeong, S. Yang, D. Zhang, S. Hu, H. Kim, J. Choi, and C. Kuo, "Perceptual quality driven frame-rate selection (PQD-FRS) for high-frame-rate video," *IEEE Transactions on Broadcasting*, vol. 62, no. 3, pp. 640–653, 2016.

[23] G. Sullivan, J. Ohm, W. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on circuits and systems for video technology*, vol. 22, no. 12, pp. 1649–1668, 2012.

[24] F. Bossen, "Common HM test conditions and software reference configurations (JCTVC-L1100)," *Joint Collaborative Team on Video Coding.*, 2013.

[25] Y. Ou, Z. Ma, T. Liu, and Y. Wang, "Perceptual quality assessment of video considering both frame rate and quantization artifacts," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 21, no. 3, pp. 286–298, 2011.

[26] F. Moss, C. Yeh, F. Zhang, R. Baddeley, and D. Bull, "Support for reduced presentation durations in subjective video quality assessment," *Signal Processing: Image Communication*, vol. 48, pp. 38–49, 2016.

[27] S. Winkler, "Analysis of public image and video databases for quality assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 616–625, 2012.

[28] M. Papadopoulos, D. Agrafiotis, and D. Bull, "On the performance of modern video coding standards with textured sequences," in *Systems, Signals and Image Processing (IWSSIP), 2015 International Conference on*. IEEE, 2015, pp. 137–140.

[29] Netflix, "Internet connection speed recommendations," https://help.netflix.com/en/node/306, [Accessed Jan. 2017].